

Adaptive Search Algorithms for Discrete Stochastic Optimization: A Smooth Best-Response Approach

Omid Namvar Gharehshiran, Vikram Krishnamurthy, and George Yin.

Abstract—This paper considers simulation-based optimization of the performance of a regime-switching stochastic system over a finite set of feasible configurations. Inspired by the stochastic fictitious play learning rules in game theory, we propose an adaptive simulation-based search algorithm that uses a smooth best-response sampling strategy and tracks the set of global optima, yet distributes the search so that most of the effort is spent on simulating the system performance at the global optima. The algorithm converges weakly to the set of global optima even when the observation data is correlated (as long as a weak law of large numbers holds). Numerical examples show that the proposed scheme yields a faster convergence for finite sample lengths compared with several existing random search and pure exploration methods in the literature.

Index Terms—Discrete stochastic optimization, Markov chain, randomized search, time-varying optima, simulation-based optimization, stochastic approximation.

I. INTRODUCTION

DISCRETE stochastic optimization problems arise in operations research [1], [2], manufacturing engineering [3], and communication networks [4], [5]. These problems are intrinsically more difficult to solve than their deterministic counterparts due to the non-availability of an explicit relation between the objective function and the underlying decision variables. It is therefore necessary to use stochastic simulation to estimate the objective function in such problems.

A. The Problem

The simplest setting of a discrete stochastic optimization problem is as follows: Estimate

$$S := \operatorname{argmin}_{s \in \mathcal{M}} F(s) = \operatorname{argmin}_{s \in \mathcal{M}} \mathbb{E} \{f_n(s)\}, \quad (1)$$

where the search space $\mathcal{M} = \{1, 2, \dots, S\}$ is finite, $\{f_n(s)\}$ for each $s \in \mathcal{M}$ is a sequence of i.i.d. random variables with finite variance but unknown distribution, \mathbb{E} denotes expectation with respect to the distribution of $f_n(s)$, and $F : \mathcal{M} \rightarrow \mathbb{R}$ is deterministic. Typically, $F(\cdot)$ represents the expected performance of a stochastic system. Since the distribution of $\{f_n(s)\}$ is unknown, $F(s)$ cannot be evaluated analytically.

A brute force method of solving (1) involves an exhaustive enumeration: For each $s \in \mathcal{M}$ compute $\hat{F}_N(s) = \frac{1}{N} \sum_{n=1}^N f_n(s)$ via simulation for large N . Then, pick $\hat{s}^* = \operatorname{argmin}_{s \in \mathcal{M}} \hat{F}_N(s)$. Since $\{f_n(s)\}$ for each $s \in \mathcal{M}$ is an i.i.d. sequence of random variables, Kolmogorov's strong law of large numbers implies that $\hat{F}_N(s) \rightarrow \mathbb{E} \{f_n(s)\}$ almost surely as $N \rightarrow \infty$. This, together with the finiteness of \mathcal{M} implies that as $N \rightarrow \infty$,

$$\operatorname{argmin}_{s \in \mathcal{M}} \hat{F}_N(s) \rightarrow \operatorname{argmin}_{s \in \mathcal{M}} \mathbb{E} \{f_n(s)\} \quad \text{w.p.1.}$$

O. N. Gharehshiran and V. Krishnamurthy are with the department of Electrical and Computer Engineering, University of British Columbia, Vancouver, V6T 1Z4, Canada (e-mail: omidn@ece.ubc.ca; vikramk@ece.ubc.ca). This research was supported by the NSERC Strategic grant and the Canada Research Chairs program.

G. Yin is with the Department of Mathematics, Wayne State University, Detroit, MI 48202, USA (e-mail: gyin@math.wayne.edu). This research was supported in part by the Army Research Office under grant W911NF-12-1-0223.

This requires $f_n(s)$ to be evaluated for each $s \in \mathcal{M}$ at each sampling period, and is highly inefficient since the evaluations $f_n(s)$, for $s \notin S$, do not contribute to finding S and are wasted. The main idea here is to develop a novel adaptive search scheme that is both *attracted* to the global optima S and *efficient*, in the sense that it spends most of its effort simulating S [6, Chapter 5.3].

Problem (1) is static in the sense that the set of global minima S does not evolve with time. In this paper, we consider two extensions of the above problem: First, we allow for $\{f_n(s)\}$ to be a correlated sequence, as long as it satisfies the weak law of large numbers. Second, we solve an adaptive variant of this problem where the set of global optima evolves with time according to the sample path of a finite-state Markov chain $\{\theta(n)\}$ with state space $\mathcal{Q} = \{1, 2, \dots, \Theta\}$. More precisely, consider a simulation-based discrete stochastic optimization problem of the form

$$S(\theta(n)) := \operatorname{argmin}_{s \in \mathcal{M}} F(s, \theta(n)) = \operatorname{argmin}_{s \in \mathcal{M}} \mathbb{E} \{f_n(s, \theta(n))\}. \quad (2)$$

We assume that the Markov chain $\{\theta(n)\}$ cannot be observed and its dynamics are unknown. However, for any choice of $s \in \mathcal{M}$, the samples $f_n(s, \theta(n))$ can be generated via simulation. We further allow time correlation of the simulation data $f_n(s, \bar{\theta})$ for each $\bar{\theta} \in \mathcal{Q}$, that is more realistic in practice.

The Markov chain $\{\theta(n)\}$ in (2) constitutes the so-called *hypermodel* [7] for the underlying dynamics. It represents the jump changes in the profile of the stochastic events in the system or the objective function or both. Such problems arise in a broad range of practical applications where the goal is to track the optimal operating configuration of a stochastic system subject to time inhomogeneity. We assume that the transition probability matrix of the Markov chain $\{\theta(n)\}$ is “close” to the identity matrix. That is, the Markov chain has transition matrix $I + \varepsilon Q$, where ε is a small parameter. We will refer to such a Markov chain with infrequent jumps as *slow Markov chain*, for simplicity. The global optima $S(\theta(n))$ thus varies with time according to the slow Markov chain. In what follows, we refer to the above problem as “regime-switching discrete stochastic optimization”. Tracking such time-varying sets lies at the very heart of applications of adaptive stochastic approximation algorithms.

Example: Consider the problem of optimizing buffer sizes in a queueing network comprising multiple stations with buffers. Such a network may represent an assembly line in the manufacturing industry, networked-processors in parallel computing, or a communication network. Let s and $\{X_n(s, \theta(n))\}$ denote the vector of buffer sizes and the sequence of random vector of service times at different stations, respectively. The distribution of service times may jump change due to the changes in the nature of the offered services. The performance of such a system $f(s, X_n(s, \theta(n)))$ is a function of both s and $\{X_n(s, \theta(n))\}$ and is often evaluated by the amortized cost of buffers minus the revenues due to the processing speed. Therefore, one seeks to minimize $F(s, \theta(n)) = \mathbb{E}_X \{f(s, X_n(s, \theta(n)))\}$ (cf. [8], [9, Chapter 2.5]).

B. Main Results

The aim is to solve the regime-switching discrete stochastic optimization problem (2). Inspired by fictitious play learning rules in game theory [10], we propose a class of adaptive search algorithms that distributes the search and evaluation functionalities efficiently. The proposed scheme can be described as follows: At each iteration n , a state $s(n)$ is sampled from the search space \mathcal{M} . The sample $s(n)$ is taken according to a randomized strategy, i.e., a probability distribution on the set \mathcal{M} , that minimizes some perturbed variant of the expected objective function based on the beliefs developed thus far. This randomized strategy is referred to *smooth best-response sampling strategy*. The perturbation term in fact simulates the search or exploration functionality essential in learning the expected stochastic behavior at various states. The objective function is then simulated at the sampled state $f_n(s(n), \theta(n))$. Finally, the simulation data is fed into a constant step-size stochastic approximation algorithm to update beliefs.

The convergence analysis in Theorem 3.1 proves that if the underlying hypermodel $\{\theta(n)\}$ evolves on the same timescale as the the proposed adaptive search scheme, the most frequently visited state tracks the set of global optima. Put differently, the algorithm spends most of its effort simulating the system at the global optima. This is desirable since, in many practical applications, the system has to be operated in the sampled configuration to measure performance. It is further shown that the proportion of time spent in non-optimal states is inversely proportional to how far their objective function values are from the global minima. The proposed algorithm relies only on the simulation data and does not require detailed information about the system model, hence, can be used directly as an on-line controller. The proposed algorithm can, as well, be deployed in static discrete stochastic optimization problems (i.e., when $\theta(n)$ is fixed); see Sec. III-D for the related discussion.

The main features of this work are:

- 1) *Correlated data*: We allow for time correlation in samples $f_n(s, \theta(n))$ that is more realistic, whereas most discrete stochastic optimization algorithms assume that the samples are i.i.d.
- 2) *Adaptive search*: The proposed algorithm tracks the optima as the underlying parameters in the discrete stochastic optimization problem evolve over time. This is in contrast to most existing algorithms that are designed to locate the optima under static settings.
- 3) *Matched timescale*: It is well known that, if the hypermodel $\theta(n)$ changes too drastically, there is no chance one can track the time-varying optima. (Such a phenomenon is known as trackability; see [7] for related discussions.) On the other hand, if $\theta(n)$ evolves on a slower timescale as compared to the adaptive search algorithm, it can be approximated by a constant on the fast timescale, hence, its variation is ignored. In this work, we consider the more difficult case where $\theta(n)$ evolves on the *same* timescale as the adaptive search algorithm and prove that the proposed scheme properly tracks the time varying optima.

Note that the proposed scheme does not assume a Markovian structure for the time-evolution of the objective function. The Markovian switching assumption is only used in our performance analysis that proceeds as follows: First, by a combined use of weak convergence methods [9] and treatment on Markov switched systems [11], [12], Theorem 4.1 in Sec. IV-A shows that the limit system for the discrete time iterates of the proposed algorithm is a randomly switching ordinary differential equation (ODE) modulated by a continuous time Markov chain. (This is in contrast to the standard treatment of stochastic approximation algorithms, where the limiting dynamics converge to a deterministic ODE.) By using multiple Lyapunov function methods for randomly switched systems [13], [14], Theorem 4.2 in Sec. IV-B proves that the limit switching ODE is asymptotically

stable almost surely. Finally, Sec. IV-D shows that tracking the global attractors set of the derived limit system provides the necessary and sufficient condition to conclude both tracking and efficiency properties of the adaptive search algorithm.

C. Literature

This work is closely connected to the literature on random search methods; see [15] for a discussion. Some random search methods spend significant effort to simulate each newly visited state at the initial stages to obtain an estimate of the objective function. Then, deploying a deterministic optimization mechanism, they search for the global optimum; see [16], [17], [18], [19]. The adaptive search algorithm in this paper is related to another class, namely, discrete stochastic approximation methods [9], [20], which distribute the simulation effort through time, and proceed cautiously based on the limited information available at each time. Algorithms from this class primarily differ in the choice of the sampling strategy. Examples of sampling strategies can be classified as : i) point-based, leading to methods such as simulated annealing [21], [22], tabu search [23], stochastic ruler [24], stochastic comparison and descent algorithms [25], [26], [27], [28], ii) set-based, leading to methods such as branch-and-bound [29], nested partitions [30], stochastic comparison and descent algorithms [31], and iii) population-based, leading to methods such as genetic algorithms.

Another related body of research pertains to the multi-armed bandit problem [32], which is concerned with optimizing the cumulative objective function values realized over a period of time, and the pure exploration problem [33], which involves finding the best arm after a given number of arm pulls. These methods seek to minimize some regret measure and, similar to the random search methods, usually assume that the problem is static in the sense that the arms' reward distributions are fixed over time¹. Further, empirical numerical studies in Sec. V reveal that bandit-based algorithms such as upper confidence bound (UCB) [32] exhibit reasonable efficiency only when the size of the search space is relatively small.

D. Organization

The rest of the paper is organized as follows: Sec. II formalizes the main assumptions posed on the problem. In Sec. III, the adaptive search scheme is presented and the main theorem of the paper entailing the tracking and efficiency properties is given. Sec. IV gives the proof of the main theorem. Finally, numerical examples are provided in Sec. V followed by the concluding remarks in Sec. VI. The proofs are relegated to the Appendix for clarity of presentation.

II. MAIN ASSUMPTIONS

This section formalizes the main assumptions posed on the regime-switching discrete stochastic optimization problem (2):

a) *Hypermodel $\theta(n)$* : A typical method for analyzing the performance of an adaptive algorithm is to postulate a hypermodel for the underlying time variations [7]. Here, we assume that all time-varying underlying parameters in the problem are finite-state and absorbed to a vector, indexed by $\theta \in \mathcal{Q}$, whose dynamics follow a discrete-time Markov chain with infrequent jumps. Condition (A1) below formally characterizes the hypermodel.

(A1) Let $\{\theta(n)\}$ be a discrete-time Markov chain with finite state space $\mathcal{Q} = \{1, 2, \dots, \Theta\}$ and transition probability matrix²

$$P^\varepsilon := I + \varepsilon Q. \quad (3)$$

¹See [34] for upper confidence bound policies for non-stationary bandit problems.

²We assume that the initial distribution of the hypermodel $\mathbf{p}_0 = [p_{0,i}]_{i \in \mathcal{Q}}$, where $P(\theta(0) = i) = p_{0,i} \geq 0$ and $\mathbf{p}_0 \mathbb{1}_\Theta = 1$, is independent of ε .

Here, $\varepsilon > 0$ is a small parameter, I denotes the $\Theta \times \Theta$ identity matrix, and $Q = [q_{ij}] \in \mathbb{R}^{\Theta \times \Theta}$ is the generator of a continuous-time Markov chain satisfying

$$q_{ij} \geq 0 \text{ for } i \neq j, |q_{ij}| \leq 1 \forall i, j \in \mathcal{Q}, Q\mathbf{1}_\Theta = \mathbf{0}, \quad (4)$$

where $\mathbf{1}_\Theta = [1, \dots, 1]_{\Theta \times 1}$ and Q is irreducible.

Choosing ε small enough ensures that the entries of P^ε in (3) are non-negative. The use of the generator Q also makes the row sum of P^ε be one. Due to the dominating identity matrix in (3), $\{\theta(n)\}$ varies slowly with time.

b) Simulation Data $f_n(s, \theta(n))$: Let \mathbb{E}_ℓ denotes the conditional expectation given \mathcal{F}_ℓ , the σ -algebra generated by $\{f_n(s, \theta(n)), s \in \mathcal{M}, \theta(n) : n < \ell\}$. We make the following assumptions.

(A2) For each $s \in \mathcal{M}$ and $\theta \in \mathcal{Q}$, $\{f_n(s, \theta)\}$ is a sequence of bounded real-valued random variables. Moreover, for any $\ell \geq 0$,

$$\frac{1}{n} \sum_{\tau=\ell}^{n+\ell-1} \mathbb{E}_\ell f_\tau(i, j) \rightarrow F(i, j) \text{ in probability as } n \rightarrow \infty, \quad (5)$$

for all $i \in \mathcal{M}$ and $j \in \mathcal{Q}$, where $F(s, \theta) = \mathbb{E}\{f_n(s, \theta)\}$; see Sec. I-A.

The above condition allows us to work with correlated processes whose remote past and distant future are asymptotically independent. Examples include: the sequence of i.i.d. random variables with (asymptotically) uniformly bounded variance, or a class of random variables (not necessarily i.i.d.) that satisfy the large deviations principle (cf. [31], [35]), e.g., moving average and stationary autoregressive processes.

Finally, we impose the following condition on the hypermodel $\theta(n)$: Let μ denote the adaptation rate of the adaptive search algorithm; see (7) or (10). Then,

(A3) $\varepsilon = \mu$ in the transition probability matrix P^ε .

Condition (A3) states that time variations of the parameters underlying the discrete stochastic optimization problem (2) occur at the same timescale as the updates in the proposed adaptive search algorithm.

Remark 2.1: It is important to stress that the hypermodel $\theta(n)$ is not used in the algorithm proposed in this paper. The algorithm does not require knowledge of $\theta(n)$ or its parameters. The hypermodel is used only in the analysis of the algorithm. We are interested in determining if the algorithm can track time-varying optima that evolve according to a slow Markov chain. Since $\theta(n)$ is unobservable, we suppress the dependence of $f_n(s, \theta(n))$ on it and, with slight abuse of notation, denote it by $f_n(s)$ in what follows.

III. TRACKING THE GLOBAL OPTIMA: ALGORITHM AND MAIN RESULTS

In this section, we introduce a stochastic approximation algorithm that, relying on smooth best-response strategies [36], [37], prescribes how to sample from the search space so as to efficiently learn and track the evolving set of global optima $\mathcal{S}(\theta(n))$. To this end, we define the smooth best-response procedure based on consecutive observations $\{f_n(s_n)\}_{n \geq 0}$ and outline its distinct properties in Sec. III-A. We then present the proposed adaptive discrete stochastic optimization algorithm in Sec. III-B followed by the main result of

the paper that shows, if one employs the proposed algorithm and the time-varying underlying parameters evolve on the same timescale as the the stochastic approximation algorithm, the algorithm efficiently tracks the set of global optima $\mathcal{S}(\theta(n))$.

A. Smooth Best-Response Sampling Strategy

Consider a learning scenario where one repeatedly samples from the search space, denoted by $s(n) \in \mathcal{M}$, at discrete times $n = 1, 2, \dots$ and obtains $f_n(s(n))$ via simulation or measurement. We postulate that $s(n)$ is chosen according to a randomized sampling strategy $\mathbf{p}(n) = (p_1(n), \dots, p_S(n))$ that belongs to the simplex of probability distributions over the search space

$$\Delta\mathcal{M} = \left\{ \mathbf{p} \in \mathbb{R}^S; p_i \geq 0, \sum_{s \in \mathcal{M}} p_i = 1 \right\}. \quad (6)$$

Based only on the collected observations $\{f_\tau(s(\tau)) : \tau \leq n\}$ up to time n , define the vector of weighted average objective function values $\tilde{\mathbf{f}}(n) = [\tilde{f}_1(n), \dots, \tilde{f}_S(n)]' \in \mathbb{R}^S$, where v' denotes the transpose of v , and

$$\tilde{f}_i(n) = \mu \sum_{\tau \leq n} (1 - \mu)^{n-\tau} \frac{f_\tau(s(\tau))}{p_i(\tau)} \cdot I_{\{s(\tau)=i\}}, \quad \forall i \in \mathcal{M}. \quad (7)$$

In (7), $I_{\{\cdot\}}$ denotes the indicator function, and the normalization factor $1/p_i(\tau)$ makes the length of the periods that each states i is chosen comparable to other states. The discount factor μ places more weight on recent observations and is necessary as the algorithm is deemed to track time-varying minima. Note further that (7) only relies on the actual measurements or simulation data $\tilde{f}_\tau(s(\tau))$ recorded (e.g. from the system performance) and does not require the system model nor the realizations of $\{\theta_\tau\}$. The smooth best-response sampling strategy is then defined as follows.

Definition 3.1: Choose a function $\rho(\sigma) : \text{int}(\Delta\mathcal{M}) \rightarrow \mathbb{R}$, where $\text{int}(G)$ denotes the interior of G and $\Delta\mathcal{M}$ is defined in (6), such that

- i) $\rho(\cdot)$ is \mathcal{C}^1 (i.e., continuously differentiable), strictly concave, and $|\rho| \leq 1$;
- ii) $\|\nabla \rho(\sigma)\| \rightarrow \infty$ as σ approaches the boundary of $\Delta\mathcal{M}$, i.e.,

$$\lim_{\sigma \rightarrow \partial(\Delta\mathcal{M})} \|\nabla \rho(\sigma)\| = \infty,$$

where $\|\cdot\|$ denotes the Euclidean norm, and $\partial(\Delta\mathcal{M})$ represents the boundary of simplex $\Delta\mathcal{M}$.

The *smooth best-response sampling* strategy is then given by

$$\mathbf{b}^\gamma(\tilde{\mathbf{f}}) := \arg \min_{\sigma \in \Delta\mathcal{M}} \sum_{i \in \mathcal{M}} \sigma_i \tilde{f}_i - \gamma \rho(\sigma), \quad 0 < \gamma < \hat{\gamma}. \quad (8)$$

The conditions imposed on the perturbation function $\rho(\cdot)$ leads to the following distinct properties of the resulting strategy:

- i) The strict concavity condition ensures the uniqueness of $\mathbf{b}^\gamma(\tilde{\mathbf{f}})$;
- ii) The boundary condition implies $\mathbf{b}^\gamma(\tilde{\mathbf{f}})$ belongs to the interior of the simplex $\Delta\mathcal{M}$.

The smooth best-response strategy is inspired by leaning algorithms in games [36], [10]. It exhibits exploration using the idea of adding a random value to the belief about the objective function values associated with each state. (This is in contrast to picking states at random with a small probability, as is common in game-theoretic learning and multi-armed bandit algorithms.) Such exploration is natural in any learning scenario. The results of [36, Theorem 2.1] show that, regardless of the distribution of the random values, a deterministic representation of the form (8) can be obtained for the pure best-response strategy resulted from adding random values to the beliefs $\tilde{\mathbf{f}}(n)$. Further, the smooth best-response strategy constructs a

³This is without loss of generality. Given an arbitrary non-absorbing Markov chain with transition probability matrix $P^\rho = I + \rho\hat{Q}$, one can form $Q = \frac{1}{q_{\max}} \cdot \hat{Q}$, where $q_{\max} = \max_{i \in \mathcal{M}} |\hat{q}_{ii}|$. To ensure that the two Markov chains generated by Q and \hat{Q} evolve on the same timescale, $\varepsilon = \rho \cdot q_{\max}$.

genuine randomized strategy. This is an appealing feature since it circumvents the discontinuity inherent in algorithms of pure best-response type (i.e., $\arg \max_{i \in \mathcal{M}} \tilde{f}_i$), where small changes in the beliefs $\tilde{\mathbf{f}}(n)$ can lead to an abrupt change in the behavior of the algorithm. Such switching behavior in the dynamics of the algorithm complicates the convergence analysis.

Remark 3.1: An example of the function $\rho(\cdot)$ in Definition 3.1 is the *entropy function* [10], [38]

$$\rho(\sigma) = - \sum_{i \in \mathcal{M}} \sigma_i \ln(\sigma_i),$$

which gives rise to the smooth best-response strategy

$$b_i^\gamma(\tilde{\mathbf{f}}) = \frac{\exp(-\tilde{f}_i/\gamma)}{\sum_{j \in \mathcal{M}} \exp(-\tilde{f}_j/\gamma)}. \quad (9)$$

Such a strategy is also used in the context of learning in games, widely known as logistic fictitious-play [39] or logit choice function [36].

B. Adaptive Discrete Stochastic Optimization Algorithm

We now proceed to present the stochastic approximation algorithm proposed for tracking the set of global optima $\mathcal{S}(\theta(n))$. The adaptive discrete stochastic optimization algorithm can be simply described as an adaptive sampling scheme. Relying on the beliefs developed about the objective function values at each states, it prescribes how to sample from the search space \mathcal{M} so as to efficiently (in terms of the amount of effort spent on simulating non-promising states) track the global optima $\mathcal{S}(\theta(n))$. We then simulate $f_n(s(n))$ at the sampled state $s(n)$ and use a stochastic approximation algorithm to update beliefs $\tilde{\mathbf{f}}(n)$ and, accordingly, the sampling strategy. The proposed algorithm relies only on the simulation data and is efficient in the sense that it requires minimum effort per iteration—it needs only one simulation, as compared to, e.g., two in [25]. Yet, as evidenced by the numerical example in Sec. V, it guarantees performance gains in terms of tracking speed.

The adaptive discrete stochastic optimization algorithm is summarized below:

Algorithm 1:

Aim. Generate a sequence $\{s(n)\}$ that provides an estimate of the time-varying global optima.

Step 0) Initialization: Choose $\rho(\cdot)$ to satisfy the conditions of Definition 3.1 and set the exploration parameter $\gamma > 0$. Initialize $\tilde{\mathbf{f}}(0) = \mathbf{0}_S$.

Step 1) State Sampling: Select state $s(n) \sim b^\gamma(\tilde{\mathbf{f}}(n))$; see (8).

Step 2) State Evaluation: Simulate or measure $f_n(s(n))$.

Step 3) Belief Update: Update the S -dimensional vector

$$\tilde{\mathbf{f}}(n+1) = \tilde{\mathbf{f}}(n) + \mu \left[\mathbf{g}(s(n), \tilde{\mathbf{f}}(n)) - \tilde{\mathbf{f}}(n) \right], \quad (10)$$

where $\mathbf{g}(s(n), \tilde{\mathbf{f}}(n))$ is a column vector with elements

$$g_i(s(n), \tilde{\mathbf{f}}(n)) := \frac{f_n(s(n))}{b_i^\gamma(\tilde{\mathbf{f}}(n))} \cdot I_{\{s(n)=i\}}. \quad (11)$$

Step 4) Recursion: Set $n \leftarrow n+1$ and go to Step 1.

Remark 3.2: 1) Note that the dynamics of $\{\theta(n)\}$ do not enter implementation of the algorithm, and is only used in the tracking analysis in Sec. IV-A. In particular, Theorem 3.1 shows that Algorithm 1 can successfully track the time-varying optima if they vary according to the hyeromodel $\theta(n)$.

2) If $\{\theta(n)\}$ was observed, one could form and update $\tilde{\mathbf{f}}_\theta(n)$ independently for each $\theta \in \mathcal{Q}$, and use $b^\gamma(\tilde{\mathbf{f}}_\theta(n))$ to select $s(n)$

once the system switched to θ' . It can then be shown that the sequence $\{s(n)\}$ spends most of its time in the global minima, irrespective of the switching, for all $\theta \in \mathcal{Q}$.

3) Larger values of γ increase the exploration weight versus exploitation, hence, decreases the time spent in $\mathcal{S}(\theta(n))$.

C. Main Result: Tracking the Regime-Switching Minima Set

To analyze the tracking capability of the above adaptive discrete stochastic optimization algorithm, define two diagnostics that will be used subsequently:

(i) *Regret* $r(n)$:

$$r(n) := \bar{F}(n) - F_{\min}(\theta(n)), \quad (12)$$

where

$$\bar{F}(n) := \mu \sum_{\tau \leq n} (1 - \mu)^{n-\tau} f_\tau(s(\tau)), \quad (13)$$

$$F_{\min}(\theta) := \min_{s \in \mathcal{M}} F(s, \theta). \quad (14)$$

Here, $\{s(n)\}$ is the sequence of states prescribed by the discrete stochastic optimization algorithm, and $\bar{F}(n)$ represents the expected realized objective function value up to sampling period n . Thus, the regret $r(n)$ quantifies the tracking capability of the algorithm.

(ii) *Empirical Sampling Distribution:* To study efficiency of the adaptive discrete stochastic optimization algorithm, we define the empirical sampling distribution vector $\mathbf{z}(n) \in \mathbb{R}^S$ as

$$\mathbf{z}(n) := \mu \sum_{\tau \leq n} (1 - \mu)^{n-\tau} \mathbf{e}_{s(\tau)}, \quad (15)$$

where $\mathbf{e}_i \in \mathbb{R}^S$ denotes the unit vector with the i th element being equal to one. Therefore, $z_i(n)$ records the percentage of iterations that state i was sampled and simulated up to time n . Efficiency of a discrete stochastic optimization algorithm is defined as the percentage of time that states within the set of global optima are sampled. For each $\bar{\theta} \in \mathcal{Q}$, the efficiency is thus quantified by $\sum_{i \in \mathcal{S}(\bar{\theta})} z_i(n)$. In (15), μ serves as the forgetting factor to facilitate adaptivity to the evolution of underlying parameters.

Before proceeding with the main theorem, define the continuous time interpolated sequence of iterates

$$\begin{aligned} \mathbf{z}^\mu(t) &= \mathbf{z}(n) \\ r^\mu(t) &= r(n) \end{aligned} \quad \text{for } t \in [n\mu, (n+1)\mu], \quad (16)$$

and let

$$\Upsilon^\eta(\theta) = \left\{ \pi \in \Delta \mathcal{M}; \sum_{i \in \mathcal{M}} \pi_i [f(i, \theta) - F_{\min}(\theta)] \leq \eta \right\}. \quad (17)$$

The following theorem asserts that the sequence $\{s(n)\}$ generated by Algorithm 1 tracks the regime-switching minima set $\mathcal{S}(\theta(n))$ and spends most of its effort on simulating $\mathcal{S}(\theta(n))$. In what follows, \Rightarrow denotes weak convergence.⁴ Note that when a sequence converges weakly to a constant, it also converges in probability to that constant.

Theorem 3.1: Suppose (A1), (A2), and (A3) hold. Let $q(\mu)$ be any sequence of real numbers satisfying $q(\mu) \rightarrow \infty$ as $\mu \rightarrow 0$. Then, for any $\eta > 0$, there exists $\bar{\gamma} > 0$ such that, if $\gamma \leq \bar{\gamma}$ in (8), as $\mu \rightarrow 0^5$:

1) *Tracking:* $(r^\mu(\cdot + q(\mu)) - \eta)^+ \Rightarrow 0$, where $x^+ = \max\{0, x\}$.

⁴Weak convergence is a generalization of convergence in distribution to a function space [9]; see also Sec. IV-A of this paper for a precise definition.

⁵We assume the initial values $\mathbf{z}(0)$ and $r(0)$ are independent of the step-size μ for simplicity. Otherwise, if $\mathbf{z}(0) = \mathbf{z}^\mu(0)$ and $r(0) = r^\mu(0)$, we require that $\mathbf{z}^\mu(0)$ and $r^\mu(0)$ converge weakly to $\mathbf{z}(0)$ and $r(0)$, respectively.

2) *Efficiency*: $\mathbf{z}^\mu(\cdot + q(\mu)) \Rightarrow \Upsilon(\theta(\cdot))$ in the sense that:

$$d(\mathbf{z}^\mu(\cdot + q(\mu)), \Upsilon^\eta(\theta(\cdot))) = \inf_{\pi(\cdot) \in \Upsilon^\eta(\theta(\cdot))} |\mathbf{z}^\mu(\cdot + q(\mu)) - \pi(\cdot)| \Rightarrow 0, \quad (18)$$

where $d(\cdot, \cdot)$ is the usual distance function, and $\theta(\cdot)$ is a continuous time Markov chain with generator Q ; see (A1).

Proof: The proof uses martingale averaging techniques to show that the limit behavior converges weakly to a switched Markovian ordinary differential equation (ODE). Then, stability of the switched ODE is established and the global attractor set is shown to represent the global minima set. The detailed proof is in Sec. IV. ■

Interpretation of Theorem 3.1: The above theorem addresses both *tracking capability* and *efficiency* of Algorithm 1: Part 1) evidences both *consistency* and *attraction* to the set $\mathcal{S}(\theta(\cdot))$ by looking at the continuous time interpolation of worst case regret $r^\mu(\cdot)$. It shows that $r^\mu(t)$ stays infinitely often less than η as $\mu \rightarrow 0$ and $t \rightarrow \infty$. (This result is similar to the *Hannan consistency* notion [40] in repeated games, however, in a regime-switching setting.) Part 2) concerns efficiency by showing that the algorithm eventually spends most of its effort on simulating $\mathcal{S}(\theta(\cdot))$ and adapts to its time variations. In particular, the proportion of time spent simulating states $s \notin \mathcal{S}(\theta(\cdot))$ is inversely proportional to how far their objective value is from the global minimum. Note that Part 2) claims convergence to a set, rather than a point in the set.

The following corollary is a direct consequence of Theorem 3.1. It asserts that the continuous time interpolation of the most frequently visited state converges weakly to the set of global minima.

Corollary 3.1: Denote the most frequently visited state by

$$s_{\max}(n) = \operatorname{argmax}_{i \in \mathcal{M}} z_i(n),$$

where $\bar{z}_i(n)$ is the i th component of $\mathbf{z}(n)$ defined in (15). Define the continuous time interpolated sequence

$$s_{\max}^\mu(t) = s_{\max}(n) \quad \text{for } t \in [n\mu, (n+1)\mu].$$

Then, under (A1)–(A3) and the conditions of Theorem 3.1, $s_{\max}^\mu(\cdot + q(\mu))$ converges weakly to the regime-switching minima set $\mathcal{S}(\theta(\cdot))$ as $\mu \rightarrow \infty$.

Note that, to foster adaptivity to the time variations of the hypermodel $\{\theta(n)\}$, Algorithm 1 selects non-optimal states with some small probability. Thus, one would not expect $\{s(n)\}$ to converge to $\mathcal{S}(\theta(n))$. In fact, $\{s(n)\}$ may visit each element of \mathcal{M} infinitely often. Instead, the strategy implemented by following Algorithm 1 ensures the empirical frequency of sampling from $\mathcal{M} \setminus \mathcal{S}(\theta(n))$ stays very low.

D. Static Discrete Stochastic Optimization

Suppose $\theta(n) = \bar{\theta}$ is fixed in (2). The discrete stochastic optimization problem then reduces to

$$\min_{s \in \mathcal{M}} F(s) = \mathbb{E} \{f_n(s, \bar{\theta})\},$$

and is *static* in the sense that the set $\mathcal{S}(\bar{\theta})$ of global minima does not evolve with time. Although not being the focus of this paper, one can use the results of [41] to show that if the exploration factor γ in (8) decreases to zero sufficiently slowly, the sequence $\{s(n)\}$ converges almost surely to $\mathcal{S}(\bar{\theta})$.

More precisely, consider the following modifications to Algorithm 1:

- (i) The constant step-size μ in (10) is replaced by decreasing step-size $\mu_n = \frac{1}{n+1}$;

- (ii) The exploration factor γ in (8) is replaced by $\frac{1}{n^\alpha}$, where $0 < \alpha < 1$.

Define the sequence of interpolated process $s^n(t)$, $n = 0, 1, \dots$:

$$\begin{aligned} s^0(t) &= s(n) \quad \text{for } t \in [t_n, t_{n+1}), \\ s^n(t) &= s^0(t + t_n) \quad \text{for } -\infty < t < \infty, \end{aligned}$$

where $t_n = \sum_{\tau=0}^{n-1} \mu_\tau$. Let $q(n)$ be any sequence of real numbers satisfying $q(n) \rightarrow \infty$ as $n \rightarrow \infty$. Then, if $\{s(n)\}$ is chosen according to Algorithm 1, $s^n(\cdot + q(n)) \xrightarrow{\text{a.s.}} \mathcal{S}(\bar{\theta})$ as $n \rightarrow \infty$ in the sense that $d(s^n(\cdot + q(n)), \mathcal{S}(\bar{\theta})) \xrightarrow{\text{a.s.}} 0$.

By the above construction, the sequence $\{s(n)\}$ will eventually become reducible with singleton communicating class $\mathcal{S}(\bar{\theta})$. That is, $\{s(n)\}$ eventually spends all its time in $\mathcal{S}(\bar{\theta})$. This is in contrast with Algorithm 1 in the regime-switching setting.

IV. PROOF OF THEOREM 3.1: TRACKING REGIME-SWITCHING GLOBAL MINIMA

This section presents the proof of the main result and is organized into four subsections: We start by showing in Sec. IV-A that the limit system associated with the discrete time iterates $(\tilde{\mathbf{f}}(n), r(n))$ is a Markovian switching system of interconnected ODEs. Next, Sec. IV-B proves that such a limit system is globally asymptotically stable with probability one and characterizes its global attractors. Accordingly, we conclude asymptotic stability of the interpolated process associated with $(\tilde{\mathbf{f}}(n), r(n))$ in Sec. IV-C, and prove that the discrete time iterates mimicking such limit dynamics is attracted to the set of global minima. Finally, Sec. IV-D uses the results obtained thus far to conclude efficiency of Algorithm 1.

A. Weak Convergence to Markovian Switching ODE

In this subsection, we use weak convergence methods to derive the limit dynamical system associated with the iterates $(\tilde{\mathbf{f}}(n), r(n))$. Before proceeding further, let us recall some definitions and notation:

Let $Z(n)$ and Z be \mathbb{R}^s -valued random vectors. We say $Z(n)$ converges weakly to Z ($Z(n) \Rightarrow Z$) if for any bounded and continuous function $\psi(\cdot)$, $E\psi(Z(n)) \rightarrow E\psi(Z)$ as $n \rightarrow \infty$. We also say that the sequence $\{Z(n)\}$ is tight if for each $\eta > 0$, there exists a compact set K_η such that $P(Z(n) \in K_\eta) \geq 1 - \eta$ for all n . The definitions of weak convergence and tightness extend to random elements in more general metric spaces. On a complete separable metric space, tightness is equivalent to relative compactness, which is known as Prohorov's Theorem [42]. By virtue of this theorem, we can extract convergent subsequences when tightness is verified. In what follows, we use a martingale problem formulation to establish the desired weak convergence. To this end, we first prove tightness. The limit process is then characterized using a certain operator related to the limit martingale problem. We refer the reader to [9, Chapter 7] for further details on weak convergence and related matters.

Define

$$\mathbf{F}(\theta) = [F(1, \theta), \dots, F(S, \theta)]', \quad (19)$$

and let

$$\hat{\mathbf{f}}(n) := \tilde{\mathbf{f}}(n) - \mathbf{F}(\theta(n)) \quad (20)$$

denote the deviation error in tracking the true objective function values via the simulation data at time n . Let further

$$\mathbf{X}(n) := \begin{bmatrix} \hat{\mathbf{f}}(n) \\ r(n) \end{bmatrix}. \quad (21)$$

It can be easily verified that $\mathbf{X}(n)$ satisfies the recursion

$$\begin{aligned} \mathbf{X}(n+1) &= \mathbf{X}(n) + \mu [\mathbf{A}_n(s(n)) - \mathbf{X}(n)] \\ &\quad + \begin{bmatrix} \mathbf{F}(\theta(n)) - \mathbf{F}(\theta(n+1)) \\ F_{\min}(\theta(n)) - F_{\min}(\theta(n+1)) \end{bmatrix}, \end{aligned} \quad (22)$$

where

$$\mathbf{A}_n(s(n)) = \begin{bmatrix} \hat{\mathbf{g}}(s(n), \hat{\mathbf{f}}(n)) - \mathbf{F}(\theta(n)) \\ f_n(s(n)) - F_{\min}(\theta(n)) \end{bmatrix}, \quad (23)$$

$$\hat{\mathbf{g}} = [\hat{g}_1, \dots, \hat{g}_S]', \quad \hat{g}_i = \frac{f_n(s(n))}{b_i^\gamma (\hat{\mathbf{f}}(n) + \mathbf{F}(\theta(n)))} \cdot I_{\{s(n)=i\}},$$

and $F_{\min}(\cdot)$ and $\mathbf{F}(\cdot)$ are defined in (14) and (19), respectively. As is widely used in the analysis of stochastic approximations, we consider the piecewise constant continuous time interpolated processes

$$\mathbf{X}^\mu(t) = \mathbf{X}(n), \quad \theta^\mu(t) = \theta(n), \quad \text{for } t \in [n\mu, (n+1)\mu]. \quad (24)$$

In what follows, we use $D([0, \infty) : \tilde{G})$ to denote the space of functions that are defined in $[0, \infty)$ taking values in \tilde{G} , and are right continuous and have left limits with Skorohod topology (see [9, p. 228]). The following theorem characterizes the limit process of the stochastic approximation iterates as a Markovian switching ODE.

Theorem 4.1: Consider the recursion (22) and suppose (A1), (A2), and (A3) hold. As $\mu \rightarrow 0$, the interpolated process $(\mathbf{X}^\mu(\cdot), \theta^\mu(\cdot))$ is tight in $D([0, \infty) : \mathbb{R}^{S+1} \times \mathcal{Q})$ and converges weakly to $(\mathbf{X}(\cdot), \theta(\cdot))$ that is a solution of the Markovian switched ODE

$$\frac{d\mathbf{X}}{dt} = \mathbf{G}(\mathbf{X}, \theta(t)) - \mathbf{X}, \quad (25)$$

where

$$\mathbf{G}(\mathbf{X}, \theta(t)) = \begin{bmatrix} \mathbf{0}_S \\ \mathbf{b}^\gamma (\hat{\mathbf{f}} + \mathbf{F}(\theta(t))) \cdot \mathbf{F}(\theta(t)) - F_{\min}(\theta(t)) \end{bmatrix}.$$

Here, $\mathbf{0}_S$ denotes an $S \times 1$ zero vector, $\mathbf{F}(\cdot)$ and $F_{\min}(\cdot)$ are defined in (14) and (19), respectively, and $\theta(t)$ denotes a continuous time Markov chain with generator Q ; see (A1).

Proof: The proof uses stochastic averaging theory based on [9]; see Appendix A for the detailed argument. ■

The above theorem asserts that the asymptotic behavior of Algorithm 1 can be captured by a dynamical system modulated by a continuous-time Markov chain $\theta(t)$. At any given instance, the Markov chain dictates which regime the system belongs to, and the system then follows the corresponding ODE until the modulating Markov chain jumps into a new state (i.e., the limit system (25) is only piecewise deterministic).

Remark 4.1: When $\theta(n)$ evolves on a slower timescale, e.g., $\varepsilon = \mathcal{O}(\mu^2)$ in (3), it remains constant in the fast timescale (i.e., the adaptive discrete stochastic optimization algorithm). Therefore, the ODE (25) will become deterministic.

B. Stability Analysis of the Markovian Switching ODE

We next proceed to analyze stability and characterize the set of global attractors of the limit system (25).

Let us start by looking at the evolution of the deviation error $\hat{\mathbf{f}}(t)$ in tracking the objective function values, which forms the first component in any trajectory $\mathbf{X}(t)$ of the limit system. In view of (25)–(26), $\hat{\mathbf{f}}(t)$ evolves according to the deterministic ODE

$$\frac{d\hat{\mathbf{f}}}{dt} = -\hat{\mathbf{f}}.$$

Note that the dynamics of $\hat{\mathbf{f}}(t)$ is independent of the second component of $\mathbf{X}(t)$, namely, the regret $r(t)$. Since the ODE is asymptotically stable, $\hat{\mathbf{f}}(t)$ decays exponentially fast to $\mathbf{0}_S$ as $t \rightarrow \infty$. This essentially establishes that realizing $f_n(s(n))$ provides sufficient

information to construct an unbiased estimator of the true objective function values⁶.

Next, substituting the global attractor $\hat{\mathbf{f}} = \mathbf{0}_S$ into the limit switching ODE associated with the regret $r(t)$ (the second component in $\mathbf{X}(t)$), we analyze stability of

$$\frac{dr}{dt} = \mathbf{b}^\gamma (\mathbf{F}(\theta(t))) \cdot \mathbf{F}(\theta(t)) - F_{\min}(\theta(t)) - r. \quad (26)$$

We start by defining stability of switched dynamical systems; see [12, Chapter 9] and [14] for further details. In what follows, $d(\cdot, \cdot)$ denotes the usual distance function.

Definition 4.1: Consider the Markovian switched system

$$\dot{Y}(t) = f(Y(t), \theta(t))$$

$$Y(0) = Y_0, \quad \theta(0) = \theta_0, \quad Y(t) \in \mathbb{R}^r, \quad \theta(t) \in \mathcal{Q},$$

where $\theta(t)$ is a continuous time Markov chain with generator Q , and $f(\cdot, i)$ is locally Lipschitz for each $i \in \mathcal{Q}$. A closed and bounded set $\mathcal{H} \subset \mathbb{R}^r \times \mathcal{Q}$ is:

- 1) *stable in probability* if for any $\varrho, \bar{\gamma} > 0$, there is a $\gamma > 0$ such that

$$\mathbb{P}\left(\sup_{t \geq 0} d((Y(t), \theta(t)), \mathcal{H}) < \bar{\gamma}\right) \geq 1 - \varrho,$$

whenever $d((Y_0, \theta_0), \mathcal{H}) < \gamma$;

- 2) *asymptotically stable in probability* if it is stable in probability and

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} d((Y(t), \theta(t)), \mathcal{H}) = 0\right) \rightarrow 1;$$

- 3) *asymptotically stable almost surely* if

$$\lim_{t \rightarrow \infty} d((Y(t), \theta(t)), \mathcal{H}) = 0 \quad \text{a.s.}$$

Before proceeding with the theorem, let

$$\mathbb{R}_{[0, \eta)} = \{r \in \mathbb{R}; 0 \leq r < \eta\}. \quad (27)$$

We break down the stability analysis of (26) into two steps; First, we examine the stability of each subsystem, i.e., for each $\bar{\theta} \in \mathcal{Q}$ when $\theta(t) = \bar{\theta}$ is fixed. The set of global attractors is shown to comprise $\mathbb{R}_{[0, \eta)}$ for all $\bar{\theta} \in \mathcal{Q}$. The slow switching condition then allows us to apply the method of multiple Lyapunov functions [43, Chapter 3] to analyze stability of the switched system.

Theorem 4.2: Consider the limit Markovian switched ODE given in (26). Let $r(0) = r_0$ and $\theta(0) = \theta_0$. For any $\eta > 0$, there exists $\bar{\gamma}(\eta)$ such that, if $\gamma < \bar{\gamma}(\eta)$ in (8), the following results hold:

- 1) If $\theta(t) = \bar{\theta}$ is fixed, the deterministic dynamical system (26) is asymptotically stable., the set $\mathbb{R}_{[0, \eta)}$ is globally asymptotically stable for each $\bar{\theta} \in \mathcal{Q}$, i.e.,

$$\lim_{t \rightarrow \infty} d(r(t), \mathbb{R}_{[0, \eta)}) = 0. \quad (28)$$

- 2) For the Markovian switching ODE, the set $\mathbb{R}_{[0, \eta)}$ is globally asymptotically stable almost surely.

Proof: For detailed proof, see Appendix B. ■

The above theorem states that the set of global attractors of the switching ODE (26) is the same as that for all non-switching ODEs (i.e., when $\theta(t) = \bar{\theta} \in \mathcal{Q}$ is fixed in (26)) and constitutes $\mathbb{R}_{[0, \eta)}$. This sets the stage for Sec. IV-D where attraction to $\mathbb{R}_{[0, \eta)}$ is shown to conclude the desired tracking and efficiency results.

⁶It can be shown that the sequence $\{\hat{\mathbf{f}}(n)\}$ induces the same asymptotic behavior as the beliefs developed using the brute force scheme [6, Chapter 5.3] about objective function values.

C. Asymptotic Stability of the Interpolated Process

In Theorem 4.1, we considered μ small and n large, but μn remained bounded. This gives a limit switched ODE for the sequence of interest as $\mu \rightarrow 0$. Here, we study asymptotic stability and establish that the limit points of the switched ODE and the stochastic approximation algorithm coincide as $t \rightarrow \infty$. We thus consider the case where $\mu \rightarrow 0$ and $n \rightarrow \infty$, however, $\mu n \rightarrow \infty$ now. Nevertheless, instead of considering a two-stage limit by first letting $\mu \rightarrow 0$ and then $t \rightarrow \infty$, we study $\mathbf{X}^\mu(t + q(\mu))$ and require $q(\mu) \rightarrow \infty$ as $\mu \rightarrow 0$. The following corollary concerns asymptotic stability of the interpolated process.

Corollary 4.1: Let

$$\mathcal{X}^\eta = \{[\mathbf{x}, r]'; \mathbf{x} = \mathbf{0}_S, r \in \mathbb{R}_{[0, \eta]}\}. \quad (29)$$

Denote by $\{q(\mu)\}$ any sequence of real numbers satisfying $q(\mu) \rightarrow \infty$ as $\mu \rightarrow 0$. Assume $\{\mathbf{X}(n) : \mu > 0, n < \infty\}$ is tight or bounded in probability. Then, for each $\eta \geq 0$, there exists $\bar{\gamma}(\eta) \geq 0$ such that if $\gamma \leq \bar{\gamma}(\eta)$ in (8),

$$\mathbf{X}^\mu(\cdot + q(\mu)) \Rightarrow \mathcal{X}^\eta, \quad \text{as } \mu \rightarrow 0. \quad (30)$$

Proof: We only give an outline of the proof, which essentially follows from Theorems 4.1 and 4.2. Define $\widehat{\mathbf{X}}^\mu(\cdot) = \mathbf{X}^\mu(\cdot + q(\mu))$. Then, it can be shown that $\widehat{\mathbf{X}}^\mu(\cdot)$ is tight. For any $T_1 < \infty$, take a weakly convergent subsequence of $\{\widehat{\mathbf{X}}^\mu(\cdot), \widehat{\mathbf{X}}^\mu(\cdot - T_1)\}$. Denote the limit by $(\widehat{\mathbf{X}}(\cdot), \widehat{\mathbf{X}}_{T_1}(\cdot))$. Note that $\widehat{\mathbf{X}}(0) = \widehat{\mathbf{X}}_{T_1}(T_1)$. The value of $\widehat{\mathbf{X}}_{T_1}(0)$ may be unknown, but the set of all possible values of $\widehat{\mathbf{X}}_{T_1}(0)$ (over all T_1 and convergent subsequences) belongs to a tight set. Using this and Theorems 4.1 and 4.2, for any $\varrho > 0$, there exists a $T_\varrho < \infty$ such that for all $T_1 > T_\varrho$, $d(\widehat{\mathbf{X}}_{T_1}(T_1), \mathcal{X}^\eta) \geq 1 - \varrho$. This implies that $d(\widehat{\mathbf{X}}(0), \mathcal{X}^\eta) \geq 1 - \varrho$, and the desired result follows. ■

D. Performance Analysis via Limit Set Characterization

The final stage of the proof deals with the analysis of efficiency and tracking properties of the adaptive discrete stochastic optimization algorithm through characterizing the limit set of the switched ODE. The result concerning the tracking capability in Theorem 3.1 follows directly from Corollary 4.1. In what follows, we use this result to conclude efficiency of Algorithm 1 by showing that the empirical sampling distribution $\mathbf{z}(n)$ tracks the set $\Upsilon^\eta(\theta(n))$ (see (17)).

Define the interpolated sequence of iterates $\bar{\mathbf{f}}^\mu(t) = \bar{\mathbf{f}}(n)$ for $t \in [n\mu, (n+1)\mu)$, and recall the interpolated processes (16). Suppose $\theta(\tau) = \bar{\theta}$ for $\tau \geq 0$. Then, in view of (13) and (15),

$$r^\mu(t) = \bar{\mathbf{f}}^\mu(t) - F_{\min}(\bar{\theta}) = \sum_{i \in \mathcal{M}} z_i^\mu(t) [f(i, \bar{\theta}) - F_{\min}(\bar{\theta})], \quad (31)$$

since $\sum_{i \in \mathcal{M}} z_i^\mu(t) = 1$. On any convergent subsequence $\{\mathbf{z}(n')\}_{n' \geq 0} \rightarrow \pi(\bar{\theta})$, with slight abuse of notation, let $\mathbf{z}^\mu(t) = \mathbf{z}(n')$ and $r^\mu(t) = r(n')$ for $t \in [n'\mu, n'\mu + \mu)$. This, together with (31), yields

$$r^\mu(\cdot + q(\mu)) \rightarrow \sum_{i \in \mathcal{M}} \pi_i(\bar{\theta}) [f(i, \bar{\theta}) - F_{\min}(\bar{\theta})], \quad \text{as } \mu \rightarrow 0, \quad (32)$$

since $q(\mu) \rightarrow 0$ as $\mu \rightarrow 0$. Finally, comparing (32) with (17) concludes that, for each $\bar{\theta} \in \mathcal{Q}$, $\mathbf{z}^\mu(\cdot + q(\mu))$ converges to the $\Upsilon^\eta(\bar{\theta})$ if and only if $r^\mu(\cdot + q(\mu)) \leq \eta$ as $\mu \rightarrow 0$. Combining this with Corollary 4.1 completes the proof of the efficiency result in Theorem 3.1.

V. NUMERICAL EXAMPLES

This section illustrates the performance of Algorithm 1 using the examples in [25], [26]. We start with a static discrete stochastic optimization example, in order to compare Algorithm 1 with two existing algorithms in the literature. We then proceed to the regime-switching framework to illustrate the tracking capability of Algorithm 1.

A. Example 1: Static Discrete Stochastic Optimization

Consider the following example described in [25, Section 4]. Suppose that the demand Y for a particular product has a Poisson distribution with parameter λ , i.e., the probability function is given by

$$d \sim f(s; \lambda) = \frac{\lambda^s \exp(-\lambda)}{s!}.$$

The objective is then to find the order size that maximizes the demand probability, subject to the constraint that at most S units can be ordered. This problem can be formulated as a discrete deterministic optimization problem:

$$\operatorname{argmax}_{s \in \{0, 1, \dots, S\}} \left[f(s; \lambda) = \frac{\lambda^s \exp(-\lambda)}{s!} \right], \quad (33)$$

which can be solved analytically. Here, we aim to solve the following stochastic variant: Compute

$$\operatorname{argmin}_{s \in \{0, 1, \dots, S\}} -\mathbb{E} \{I_{\{d=s\}}\}, \quad (34)$$

where $I_{\{\cdot\}}$ denotes the indicator function, and d is a Poisson distributed random variable with rate λ . Clearly, problems (33) and (34) both lead to the same set of global optimizers. This enables us to check the results obtained using Algorithm 1.

We consider the following two cases of the rate parameter λ in (33): i) $\lambda = 1$, which implies that the set of global optimizers is $\mathcal{S} = \{0, 1\}$, and ii) $\lambda = 10$, in which case the set of global optimizers is $\mathcal{S} = \{9, 10\}$. For each case, we further study the effect of the search space size on the performance of Algorithm 1 by considering two instances: i) $S = 10$, and ii) $S = 100$. Finally, we compare Algorithm 1 (referred to as AS) with the following two algorithms that have been proposed in the literature:

- i) Random search (RS) [25]: Each iteration of the RS algorithm requires one random number selection, $\mathcal{O}(S)$ arithmetic operations, one comparison and two independent evaluations of the objective function $f_n(s)$.
- ii) Upper confidence bound (UCB) [32]: Each iteration of the UCB algorithm requires $\mathcal{O}(S)$ arithmetic operations, one maximization and one evaluation of the objective function $f_n(s)$.

Note in comparison that, using $\rho(x)$ as in Remark 3.1, the AS algorithm proposed in this paper requires $\mathcal{O}(S)$ arithmetic operations, one random number selection and one evaluation of the objective function $f_n(s)$ at each iteration. Since the problem is static in the sense that \mathcal{S} is fixed for each case, we apply the modifications discussed in Sec. III-D to Algorithm 1 and set $\alpha = 0.2$ and $\gamma = 0.01$ in this example.

To give a fair comparison of the three algorithms, we use the iteration number to denote the number of performed simulations. All three algorithms are initialized at state $s(0)$, that is chosen uniformly from \mathcal{M} , and move towards \mathcal{S} . Close scrutiny of the results presented in Table I leads to the following observations: In all three algorithms, the speed of convergence decreases when either S or λ (or both) increases. However, the effect of increasing λ is more substantial since the objective function values of the worst and best states become closer when $\lambda = 10$. At a fixed iteration number, higher percentage of cases where a particular method has converged to the global optima indicates convergence at a faster rate. As the results of Table I show,

TABLE I

EXAMPLE 1: PERCENTAGE OF CASES WHERE ALGORITHMS CONVERGED TO GLOBAL OPTIMA $\mathcal{S}(\theta(n))$ IN n ITERATIONS

RS: Random Search Algorithm of [25]
 AS: Proposed Adaptive Search in Algorithm 1
 UCB: Upper Confidence Bound Algorithm of [32]
 (a) $\lambda = 1$

| Iteration n | $S = 10$ | | | $S = 100$ | | |
|------------------|----------|-----|-----|-----------|-----|-----|
| | AS | RS | UCB | AS | RS | UCB |
| 10 | 55 | 39 | 86 | 11 | 6 | 43 |
| 50 | 98 | 72 | 90 | 30 | 18 | 79 |
| 100 | 100 | 82 | 95 | 48 | 29 | 83 |
| 500 | 100 | 96 | 100 | 79 | 66 | 89 |
| 1000 | 100 | 100 | 100 | 93 | 80 | 91 |
| 5000 | 100 | 100 | 100 | 100 | 96 | 99 |
| 10000 | 100 | 100 | 100 | 100 | 100 | 100 |

(b) $\lambda = 10$

| Iteration n | $S = 10$ | | | $S = 100$ | | |
|------------------|----------|----|-----|-----------|----|-----|
| | AS | RS | UCB | AS | RS | UCB |
| 10 | 29 | 14 | 15 | 7 | 3 | 2 |
| 100 | 45 | 30 | 41 | 16 | 9 | 13 |
| 500 | 54 | 43 | 58 | 28 | 21 | 25 |
| 1000 | 69 | 59 | 74 | 34 | 26 | 30 |
| 5000 | 86 | 75 | 86 | 60 | 44 | 44 |
| 10000 | 94 | 84 | 94 | 68 | 49 | 59 |
| 20000 | 100 | 88 | 100 | 81 | 61 | 74 |
| 50000 | 100 | 95 | 100 | 90 | 65 | 81 |

Algorithm 1 ensures faster convergence to the global optima \mathcal{S} in each case.

To illustrate superior efficiency of Algorithm 1, we plot the sample path of the time spent simulating states outside the global optima, i.e.,

$$1 - \sum_{i \in \mathcal{S}} z_i(n), \quad (35)$$

in Fig. 1. This figure corresponds to the case where $\lambda = 1$ and $S = 100$ in (34). As can be seen, since the RS method randomizes among all states (except the previously sampled state) at each iteration, it spends roughly 98% of its simulation effort on non-optimal states. Further, the UCB algorithm switches to its exploitation phase after a longer period of exploration as compared to Algorithm 1. Fig 1 thus indicates that Algorithm 1 guarantees a superior balance between exploration of the search space and exploitation of the collected data as compared to other schemes.

B. Example 2: Regime-Switching Discrete Stochastic Optimization

Consider the discrete stochastic optimization problem described in Example 1 with the exception that now $\lambda(\theta(n))$ jump changes between 1 and 10 according to a slow Markov chain $\{\theta(n)\}$ with state space $\mathcal{Q} = \{1, 2\}$, and transition probability matrix

$$P^\varepsilon = I + \varepsilon Q, \quad Q = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}. \quad (36)$$

More precisely, $\lambda(1) = 1$ and $\lambda(2) = 10$. Assuming $S = 10$, $\mathcal{S}(1) = \{0, 1\}$ and $\mathcal{S}(2) = \{9, 10\}$. Then, the discrete stochastic optimization problem is given by (34), where

$$d \sim f(s, i; \lambda) = \frac{\lambda^s(i) \exp(-\lambda(i))}{s!}, \quad i = 1, 2. \quad (37)$$

In the rest of this section, we assume $\gamma = 0.1$, and $\mu = \varepsilon = 0.01$. Further, we shall use an adaptive variant of RS, studied in [27], and

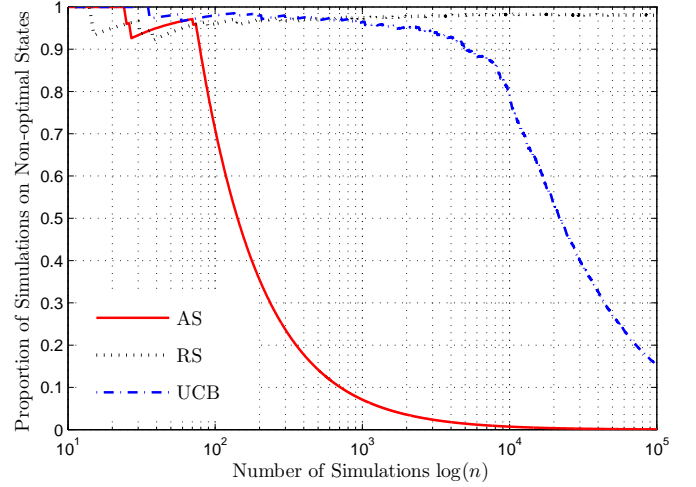


Fig. 1. Example 1: Proportion of simulation effort expended on states outside the global optima set ($\lambda = 1$, $S = 100$).

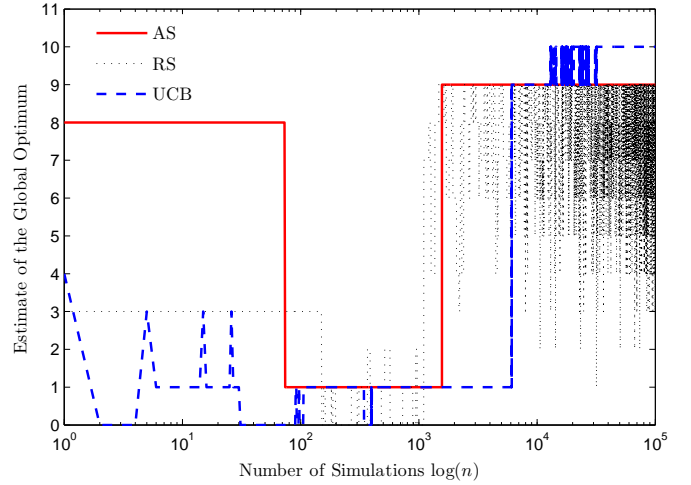


Fig. 2. Example 2: Sample path of the estimate of the global optima when $S = 10$ and the global optima evolve with time. For $0 \leq n < 10^3$, the global optima set is $\{0, 1\}$. For $10^3 \leq n < 10^5$, the global optima set is $\{9, 10\}$.

an adaptive variant of UCB both with constant step-sizes $\mu = 0.01$ to compare with algorithm 1.

Fig. 2 shows tracking capability of Algorithm 1 when the Markov chain $\{\theta(n)\}$ undergoes a jump from $\theta = 1$ to $\theta = 2$ at $n = 10^3$. As can be seen, contrary to the RS algorithm, both AS and UCB methods properly track the changes; however, AS is more agile. Superior performance of the AS algorithm is further verified in Fig. 3 which shows how the simulation effort on non-optimal states evolves as the rate parameter λ jump changes. Fig. 3 thus confirms that the superior balance between exploration and exploitation properly responds to the regime switching.

Fig. 4 illustrates the efficiency (35) of the AS algorithm for several values of ε . Note that ε represents the speed of Markovian switching. Each point on the graph is an average over 100 independent runs of 10^6 iterations of the algorithms when (36) is adopted as the transition matrix of $\{\theta(n)\}$. As expected, the percentage of samples taken from the set of global optima increases for all methods as the speed of time variations decreases; however, superior efficiency of the AS algorithm is clearly evident.

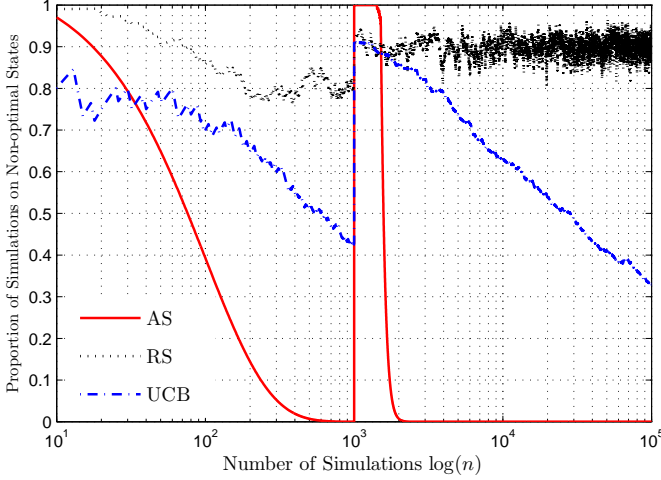


Fig. 3. Example 2: Proportion of simulation effort expended on non-optimal states when $S = 10$ and the global optima evolve with time. For $0 \leq n < 10^3$, the global optima set is $\{0, 1\}$. For $10^3 \leq n < 10^5$, the global optima set is $\{9, 10\}$.

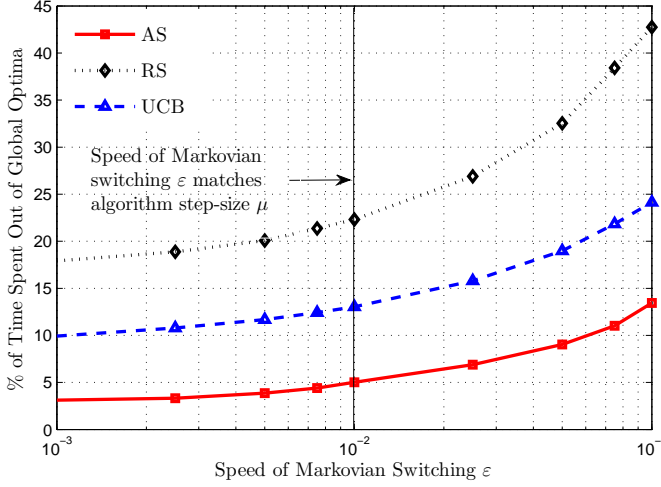


Fig. 4. Example 2: Proportion of time the estimate of global optima spends out of the global optima set $\mathcal{S}(\theta(n))$ versus the speed of Markovian switching of the set of global optimizers ($S = 10$).

VI. CONCLUSION

This paper has considered regime-switching discrete stochastic optimization problems where the underlying time variations, e.g., in the profile of the stochastic behavior of the system or the objective function, can be captured by the sample path of a slow discrete time Markov chain. We proposed a class of adaptive search algorithms that prescribes how to iteratively sample states from the search space. The proposed scheme is a constant step-size stochastic approximation algorithm that updates beliefs about the objective function values, accompanied by an adaptive sampling strategy of best-response type. The convergence analysis proved that, if the underlying time vari-

ations occur on the same timescale as the stochastic approximation algorithm, the algorithm will properly track the randomly switching set of global minima. Further, the proposed scheme ensures “most” of the simulation effort is spent on the global minima. It thus can be deployed as an on-line control mechanism to enable self-configuration of large scale stochastic systems. The main features of the proposed adaptive discrete stochastic optimization algorithm include: 1) it allows time correlation in the sampled data; 2) it tracks time varying optima when the parameters underlying the stochastic optimization problem evolve over time; 3) in contrast to the case where the time variations occur on a slower timescale as the adaptive search algorithm and trackability is trivial, it tracks time variations of the global optima even when such variations occur on the same timescale as the updates of the proposed algorithm. Numerical examples illustrated the trade-off between efficiency (the number of executed simulations) and the convergence speed, as compared with the existing random search and pure exploration methods.

APPENDIX A

PROOF OF THEOREM 4.1

We first prove tightness of the interpolated process $\mathbf{X}^\mu(\cdot)$. Consider the sequence $\{\mathbf{X}(n)\}$, defined in (21). In view of the boundedness of the objective function, and by virtue of Hölder’s and Gronwall’s inequalities, for any $0 < T_1 < \infty$,

$$\sup_{k \leq T_1/\mu} \mathbb{E} \|\mathbf{X}(k)\|^2 < \infty, \quad (38)$$

where in the above and hereafter $\|\cdot\|$ denotes the Euclidean norm and t/μ is understood to be the integer part of t/μ for each $t > 0$. Next, considering the interpolated process $\mathbf{X}^\mu(\cdot)$ (defined in (24)) and the recursion (22), for any $t, u > 0$, $\delta > 0$, and $u < \delta$, it can be verified that

$$\begin{aligned} \mathbf{X}^\mu(t+u) - \mathbf{X}^\mu(t) &= \mu \sum_{k=t/\mu}^{(t+u)/\mu-1} [\mathbf{A}_k(s(k)) - \mathbf{X}(k)] \\ &+ \sum_{k=t/\mu}^{(t+u)/\mu-1} \left[\frac{\mathbf{F}(\theta(k)) - \mathbf{F}(\theta(k+1))}{F_{\min}(\theta(k)) - F_{\min}(\theta(k+1))} \right], \end{aligned} \quad (39)$$

where $\mathbf{A}_k(s(k))$ is defined in (23). Consequently, using the parallelogram law,

$$\begin{aligned} \mathbb{E}_t^\mu \|\mathbf{X}^\mu(t+u) - \mathbf{X}^\mu(t)\|^2 &\leq 2\mathbb{E}_t^\mu \left\| \mu \sum_{k=t/\mu}^{(t+u)/\mu-1} \mathbf{A}_k(s(k)) - \mathbf{X}(k) \right\|^2 \\ &+ 2\mathbb{E}_t^\mu \left\| \sum_{k=t/\mu}^{(t+u)/\mu-1} \left[\frac{\mathbf{F}(\theta(k)) - \mathbf{F}(\theta(k+1))}{F_{\min}(\theta(k)) - F_{\min}(\theta(k+1))} \right] \right\|^2, \end{aligned} \quad (40)$$

where \mathbb{E}_t^μ denotes the σ -algebra generated by the μ -dependent past data up to time t . By virtue of the tightness criteria [44, Theorem 3, p. 47] or [9, Chapter 7], it suffices to verify

$$\lim_{\delta \rightarrow 0} \limsup_{\mu \rightarrow 0} \left\{ \mathbb{E} \left[\sup_{0 \leq u \leq \delta} \mathbb{E}_t^\mu \|\mathbf{X}^\mu(t+u) - \mathbf{X}^\mu(t)\|^2 \right] \right\} = 0. \quad (41)$$

As for the first term on the r.h.s. of (40), noting the boundedness of objective function, we obtain: see (42) at the bottom of the page.

$$\begin{aligned} \mathbb{E}_t^\mu \left\| \mu \sum_{k=t/\mu}^{(t+u)/\mu-1} \mathbf{A}_k(s(k)) - \mathbf{X}(k) \right\|^2 &= \mu^2 \sum_{\tau=t/\mu}^{(t+u)/\mu-1} \sum_{\kappa=t/\mu}^{(t+u)/\mu-1} \mathbb{E}_t^\mu \{ [\mathbf{A}_\tau(s(\tau)) - \mathbf{X}(\tau)]' [\mathbf{A}_\kappa(s(\kappa)) - \mathbf{X}(\kappa)] \} \\ &\leq K\mu^2 \left(\frac{t+u}{\mu} - \frac{t}{\mu} \right)^2 = O(u^2) \end{aligned} \quad (42)$$

We then concentrate on the second term on the r.h.s. of (40). Note that, for sufficiently small positive μ , if Q is irreducible, then so is $I + \mu Q$. Thus, for sufficiently large k , $\|(I + \mu Q)^k - \mathbb{1}\nu_\mu\|_M \leq \lambda_c^k$ for some $0 < \lambda_c < 1$, where ν_μ denotes the row vector of stationary distribution associated with the transition matrix $I + \mu Q$, $\mathbb{1}$ denotes the column vector of ones, and $\|\cdot\|_M$ represents any matrix norm. The essential feature involved in the second term in (40) is the difference of the transition probability matrix of the form $(I + \mu Q)^{k-(t/\mu)} - (I + \mu Q)^{k+1-(t/\mu)}$. However, it can be seen that

$$\begin{aligned} & (I + \mu Q)^{k-(t/\mu)} - (I + \mu Q)^{k+1-(t/\mu)} \\ &= -\mu Q[(I + \mu Q)^{k-(t/\mu)} - \mathbb{1}\nu_\mu]. \end{aligned}$$

This in turn implies that

$$\begin{aligned} & \mathbb{E}_t^\mu \left\| \sum_{k=t/\mu}^{(t+u)/\mu-1} \begin{bmatrix} \mathbf{F}(\theta(k)) - \mathbf{F}(\theta(k+1)) \\ F_{\min}(\theta(k)) - F_{\min}(\theta(k+1)) \end{bmatrix} \right\|^2 \\ & \leq KO(\mu) \sum_{k=t/\mu}^{(t+u)/\mu-1} \lambda_c^k \leq O(\mu) \sum_{k=1}^{\infty} \lambda_c^k = O(\mu). \end{aligned} \quad (43)$$

Finally, combining (42) and (43), the tightness criteria (41) is verified. Therefore, $\mathbf{X}^\mu(\cdot)$ is tight in $D([0, \infty] : \mathbb{R}^{S+1})$. In view of [27, Proposition 4.4], $\theta^\mu(\cdot)$ is also tight and $\theta^\mu(\cdot) \Rightarrow \theta(\cdot)$ such that $\theta(\cdot)$ is a continuous time Markov chain with generator Q ; see (A1). As the result, the pair $(\mathbf{X}^\mu(\cdot), \theta^\mu(\cdot))$ is tight in $D([0, \infty] : \mathbb{R}^{S+1} \times \mathcal{Q})$.

Using Prohorov's theorem [9], one can extract a convergent subsequence. For notational simplicity, we still denote the subsequence by $\mathbf{X}^\mu(\cdot)$ with limit $\mathbf{X}(\cdot)$. By the Skorohod representation theorem [9], and with a slight abuse of notation, $\mathbf{X}^\mu(\cdot) \rightarrow \mathbf{X}(\cdot)$ in the sense of w.p.1 and the convergence is uniform on any compact interval. We now proceed to characterize the limit $\mathbf{X}(\cdot)$ using martingale averaging methods.

First, we demonstrate that the last term in (39) contributes nothing to the limit differential equation. We aim to show

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}_t h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times E_t^\mu \left[\sum_{k=t/\mu}^{(t+u)/\mu-1} \begin{bmatrix} \mathbf{F}(\theta(k)) - \mathbf{F}(\theta(k+1)) \\ F_{\min}(\theta(k)) - F_{\min}(\theta(k+1)) \end{bmatrix} \right] = 0. \end{aligned}$$

This directly follows from an argument similar to the one used in (43).

To obtain the desired limit, it will be proved that the limit $(\mathbf{X}(\cdot), \theta(\cdot))$ is the solution of the martingale problem with operator \mathcal{L} defined as follows: For all $i \in \mathcal{Q}$,

$$\begin{aligned} \mathcal{L}y(x, i) &= \nabla_x y'(x, i) [\mathbf{G}(x, i) - x] + Qy(x, \cdot)(i), \\ Qy(x, \cdot)(i) &= \sum_{j \in \mathcal{Q}} q_{ij} y(x, j), \end{aligned} \quad (44)$$

and, for each $i \in \mathcal{Q}$, $y(\cdot, i) : \mathbb{R}^r \mapsto \mathbb{R}$ with $y(\cdot, i) \in C_0^1$ (C^1 function with compact support). Further, $\nabla_x y(x, i)$ denotes the gradient of $y(x, i)$ with respect to x , and $\mathbf{G}(\cdot, \cdot)$ is defined in (26). Using an argument similar to [11, Lemma 7.18], one can show that the martingale problem associated with the operator \mathcal{L} has a unique solution. Thus, it remains to prove that the limit $(\mathbf{X}(\cdot), \theta(\cdot))$ is the solution of the martingale problem. To this end, it suffices to show that, for any positive arbitrary integer κ_0 , and for any $t, u > 0$, $0 < t_\iota \leq t$ for all $\iota \leq \kappa_0$, and any bounded continuous function $h(\cdot, i)$ for all $i \in \mathcal{Q}$,

$$\begin{aligned} & \mathbb{E} h(\mathbf{X}(t_\iota), \theta(t_\iota) : \iota \leq \kappa_0) \\ & \times \left[y(\mathbf{X}(t+u), \theta(t+u)) - y(\mathbf{X}(t), \theta(t)) \right. \\ & \left. - \int_t^{t+u} \mathcal{L}y(\mathbf{X}(v), \theta(v)) dv \right] = 0. \end{aligned} \quad (45)$$

To verify (45), we work with $(\mathbf{X}^\mu(\cdot), \theta^\mu(\cdot))$ and prove that the above equation holds as $\mu \rightarrow 0$.

By the weak convergence of $(\mathbf{X}^\mu(\cdot), \theta^\mu(\cdot))$ to $(\mathbf{X}(\cdot), \theta(\cdot))$ and Skorohod representation, it can be seen that

$$\begin{aligned} & \mathbb{E} h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times [(\mathbf{X}^\mu(t+u), \theta^\mu(t+u)) - (\mathbf{X}^\mu(t), \theta^\mu(t))] \\ & \rightarrow \mathbb{E} h(\mathbf{X}(t_\iota), \theta(t_\iota) : \iota \leq \kappa_0) \\ & \times [(\mathbf{X}(t+u), \theta(t+u)) - (\mathbf{X}(t), \theta(t))]. \end{aligned}$$

Now, choose a sequence of integers $\{n_\mu\}$ such that $n_\mu \rightarrow \infty$ as $\mu \rightarrow 0$, but $\delta_\mu = \mu n_\mu \rightarrow 0$, and Partition $[t, t+u]$ into subintervals of length δ_μ . Then,

$$\begin{aligned} & y(\mathbf{X}^\mu(t+u), \theta^\mu(t+u)) - y(\mathbf{X}^\mu(t), \theta^\mu(t)) \\ &= \sum_{\ell: \ell\delta_\mu=t}^{t+u} \left[y(\mathbf{X}(\ell n_\mu + n_\mu), \theta(\ell n_\mu + n_\mu)) \right. \\ & \quad \left. - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) \right] \\ &= \sum_{\ell: \ell\delta_\mu=t}^{t+u} \left[y(\mathbf{X}(\ell n_\mu + n_\mu), \theta(\ell n_\mu + n_\mu)) \right. \\ & \quad \left. - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu + n_\mu)) \right] \\ &+ \sum_{\ell: \ell\delta_\mu=t}^{t+u} \left[y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu + n_\mu)) \right. \\ & \quad \left. - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) \right], \end{aligned} \quad (46)$$

where $\sum_{\ell: \ell\delta_\mu=t}^{t+u}$ denotes the sum over ℓ in the range $t \leq \ell\delta_\mu \leq t+u$.

First, we consider the second term on the r.h.s. of (46): see (47) at the bottom of the next page. As for the first term on the r.h.s. of (46): see (48) at the bottom of the next page, where $\nabla_x y$ denotes the gradient column vector with respect to vector x , $\nabla_x' y$ represents its transpose, and $\hat{\mathbf{g}}_\theta(\cdot, \cdot)$ denotes the vector $\hat{\mathbf{g}}(\cdot, \cdot)$ in (23) when $\theta(k) = \theta$ is held fixed. The rest of the proof is divided into two steps, each concerning one of the two terms in (48). For notational simplicity, we shall write $\nabla_{\hat{\mathbf{f}}} y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu))$, and $\nabla_r y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu))$ as $\nabla_{\hat{\mathbf{f}}} y$, and $\nabla_r y$, respectively.

Step 1: We start by looking at

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E} h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla_{\hat{\mathbf{f}}} y \right. \\ & \quad \times \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \left[\hat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \hat{\mathbf{f}}(k)) - \mathbf{F}(\theta(\ell n_\mu)) \right] \right] \\ &= \lim_{\mu \rightarrow 0} \mathbb{E} h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla_{\hat{\mathbf{f}}} y \right. \\ & \quad \times \left[-\mathbf{F}(\theta(\ell n_\mu)) + \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \sum_{\hat{\theta}=1}^{\Theta} \right. \\ & \quad \left. \left[\sum_{\theta=1}^{\Theta} \mathbb{E}_{\ell n_\mu} \hat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \hat{\mathbf{f}}(k)) \mathbb{E}_{\ell n_\mu} I_{\{\theta(k)=\theta | \theta(\ell n_\mu)=\hat{\theta}\}} \right] \right] \end{aligned} \quad (49)$$

We concentrate on the term involving the Markov chain $\theta(k)$. Note that for large k with $\ell n_\mu \leq k \leq \ell n_\mu + n_\mu$ and $k - \ell n_\mu \rightarrow \infty$, by [27, Proposition 4.4], for some $\hat{k}_0 > 0$,

$$\begin{aligned} & (I + \mu Q)^{k - \ell n_\mu} = Z((k - \ell n_\mu)\mu) \\ & \quad + O(\mu + \exp(-\hat{k}_0(k - \ell n_\mu))), \end{aligned}$$

$$\frac{dZ(t)}{dt} = Z(t)Q, \quad Z(0) = I.$$

For $\ell n_\mu \leq k \leq \ell n_\mu + n_\mu$, letting $\mu \ell n_\mu \rightarrow u$ yields that $(k - \ell n_\mu)\mu \rightarrow 0$ as $\mu \rightarrow 0$. For such k , $Z((k - \ell n_\mu)\mu) \rightarrow I$. Therefore, by the boundedness of $\widehat{\mathbf{g}}(s(k), \widehat{\mathbf{f}}(k))$, it follows that, as $\mu \rightarrow 0$,

$$\begin{aligned} & \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \left\| \mathbb{E}_{\ell n_\mu} \widehat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \widehat{\mathbf{f}}(k)) \right\| \\ & \times \left| \mathbb{E}_{\ell n_\mu} \left[I_{\{\theta(k)=\theta\}} I_{\{\theta(\ell n_\mu)=\check{\theta}\}} \right] - I_{\{\theta(\ell n_\mu)=\check{\theta}\}} \right| \rightarrow 0. \end{aligned}$$

Therefore,

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \delta_\mu \nabla'_{\widehat{\mathbf{f}}} y \right. \\ & \times \left. \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \left[\widehat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \widehat{\mathbf{f}}(k)) - \mathbf{F}(\theta(\ell n_\mu)) \right] \right] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \delta_\mu \nabla'_{\widehat{\mathbf{f}}} y \sum_{\check{\theta}=1}^{\Theta} I_{\{\theta(\ell n_\mu)=\check{\theta}\}} \right. \\ & \times \left. \left[-F(\check{\theta}) + \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \mathbb{E}_{\ell n_\mu} \widehat{\mathbf{g}}_{\check{\theta}}(s(k), \widehat{\mathbf{f}}(k)) \right] \right]. \end{aligned} \quad (50)$$

It is more convenient to work with the individual elements of $\widehat{\mathbf{g}}_{\check{\theta}}(\cdot, \cdot)$. Substituting for the i -th element from (23) in (50) results

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \delta_\mu \nabla'_{\widehat{\mathbf{f}}} y \sum_{\check{\theta}=1}^{\Theta} I_{\{\theta(\ell n_\mu)=\check{\theta}\}} \left[-F(i, \check{\theta}) \right. \right. \\ & \left. \left. + \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \mathbb{E}_{\ell n_\mu} \left\{ \frac{f_k(i)}{b_i^\gamma(\widehat{\mathbf{f}}(k) + \mathbf{F}(\check{\theta}))} \cdot I_{\{s(k)=i\}} \right\} \right] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \delta_\mu \nabla'_{\widehat{\mathbf{f}}} y \right. \\ & \times \sum_{\check{\theta}=1}^{\Theta} \left[-F(i, \check{\theta}) + \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \mathbb{E}_{\ell n_\mu} f_k(i) \right] I_{\{\theta(\ell n_\mu)=\check{\theta}\}} \left. \right]. \end{aligned} \quad (51)$$

In (51), we used $\mathbb{E}_{\ell n_\mu} I_{\{s(k)=i\}} = b_i^\gamma(\widehat{\mathbf{f}}(\ell n_\mu))$ since $s(k)$ is chosen according to the smooth best-response strategy $\mathbf{b}^\gamma(\widehat{\mathbf{f}}(k))$; see Step 1) in Algorithm 1. Note that $f_k(i)$ is still time-dependent due to the presence of noise in the simulation data. Note further that $\theta(\ell n_\mu) = \theta^\mu(\mu \ell n_\mu)$. In light of (C1)–(C2), by the weak convergence of $\theta^\mu(\cdot)$ to $\theta(\cdot)$, the Skorohod representation, and using $\mu \ell n_\mu \rightarrow u$, it can be shown for the second term in (51) that, as $\mu \rightarrow 0$,

$$\begin{aligned} & \sum_{\check{\theta}=1}^{\Theta} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \mathbb{E}_{\ell n_\mu} f_k(i) I_{\{\theta^\mu(\mu \ell n_\mu)=\check{\theta}\}} \\ & \rightarrow \sum_{\check{\theta}=1}^{\Theta} F(i, \check{\theta}) I_{\{\theta(u)=\check{\theta}\}} = F(i, \theta(u)) \quad \text{in probability.} \end{aligned} \quad (52)$$

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} [y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu + \mu)) - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu))] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \sum_{i_0=1}^{\Theta} \sum_{j_0=1}^{\Theta} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} [y(\mathbf{X}(\ell n_\mu), j_0) \mathbb{P}(\theta(k+1) = j_0 | \theta(k) = i_0) - y(\mathbf{X}(\ell n_\mu), i_0)] I_{\{\theta(k)=i_0\}} \right] \\ & = \mathbb{E}h(\mathbf{X}(t_\iota), \theta(t_\iota) : \iota \leq \kappa_0) \left[\int_t^{t+u} Qy(\mathbf{X}(v), \theta(v)) dv \right] \end{aligned} \quad (47)$$

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} [y(\mathbf{X}(\ell n_\mu + n_\mu), \theta(\ell n_\mu + n_\mu)) - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu + n_\mu))] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} [y(\mathbf{X}(\ell n_\mu + n_\mu), \theta(\ell n_\mu)) - y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu))] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \nabla'_{\widehat{\mathbf{f}}} y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) [\widehat{\mathbf{f}}(\ell n_\mu + n_\mu) - \widehat{\mathbf{f}}(\ell n_\mu)] \right. \\ & \quad \left. + \nabla'_r y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) [r(\ell n_\mu + n_\mu) - r(\ell n_\mu)] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \times \left[\sum_{\ell: \ell \delta_\mu = t}^{t+u} \delta_\mu \nabla'_{\widehat{\mathbf{f}}} y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} [\widehat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \widehat{\mathbf{f}}(k)) - \mathbf{F}(\theta(\ell n_\mu))] - \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \widehat{\mathbf{f}}(k) \right] \right. \\ & \quad \left. + \delta_\mu \nabla'_r y(\mathbf{X}(\ell n_\mu), \theta(\ell n_\mu)) \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} [f_k(s(k)) - F_{\min}(\theta(\ell n_\mu))] - \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} r(k) \right] \right] \end{aligned} \quad (48)$$

Using a similar argument for the first term in (51) yields

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla'_{\tilde{\mathbf{f}}} y \right. \\ & \quad \times \left. \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} [\hat{\mathbf{g}}_{\theta(\ell n_\mu)}(s(k), \hat{\mathbf{f}}(k)) - \mathbf{F}(\theta(\ell n_\mu))] \right] \right] \\ & \rightarrow \mathbf{0}_S \quad \text{as } \mu \rightarrow 0. \end{aligned} \quad (53)$$

By using the technique of stochastic approximation (see, e.g., [9, Chapter 8]), it can be shown that

$$\begin{aligned} & \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla'_{\tilde{\mathbf{f}}} y \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \hat{\mathbf{f}}(k) \right] \right] \\ & \rightarrow \mathbb{E}h(\mathbf{X}(t_\iota), \theta(t_\iota) : \iota \leq \kappa_0) \\ & \quad \times \left[\int_t^{t+u} \nabla'_{\tilde{\mathbf{f}}} y(\mathbf{X}(v), \theta(v)) \hat{\mathbf{f}}(v) dv \right] \quad \text{as } \mu \rightarrow 0. \end{aligned} \quad (54)$$

Step 2: Next, we concentrate on the second term in (48). By virtue of the boundedness of $f_k(s(k))$, and using a similar argument as in Step 1,

$$\begin{aligned} & \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \frac{\delta_\mu}{n_\mu} \nabla'_r y \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \right. \\ & \quad \left. \left[\sum_{\tilde{\theta}=1}^{\Theta} \left[\sum_{i=1}^S \mathbb{E}_{\ell n_\mu} f_k(i) I_{\{s(k)=i\}} - F_{\min}(\tilde{\theta}) \right] I_{\{\theta(\ell n_\mu)=\tilde{\theta}\}} \right] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \frac{\delta_\mu}{n_\mu} \nabla'_r y \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \right. \\ & \quad \left. \left[\sum_{\tilde{\theta}=1}^{\Theta} \left[\sum_{i=1}^S b_i^\gamma(\tilde{\mathbf{f}}(\ell n_\mu)) \mathbb{E}_{\ell n_\mu} f_k(i) - F_{\min}(\tilde{\theta}) \right] I_{\{\theta(\ell n_\mu)=\tilde{\theta}\}} \right] \right] \\ & = \lim_{\mu \rightarrow 0} \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla'_r y \right. \\ & \quad \times \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} \sum_{\tilde{\theta}=1}^{\Theta} I_{\{\theta^\mu(\mu \ell n_\mu)=\tilde{\theta}\}} \right. \\ & \quad \times \left. \left[\sum_{i=1}^S b_i^\gamma(\tilde{\mathbf{f}}^\mu(\mu \ell n_\mu)) \mathbb{E}_{\ell n_\mu} f_k(i) - F_{\min}(\tilde{\theta}) \right] \right] \right]. \end{aligned} \quad (55)$$

Here, we used $\mathbb{E}_{\ell n_\mu} I_{\{s(k)=i\}} = b_i^\gamma(\tilde{\mathbf{f}}(\ell n_\mu))$ as in Step 1. Recall that $\tilde{\mathbf{f}}(k) = \hat{\mathbf{f}}(k) + \mathbf{F}(\theta(k))$; see (20). By weak convergence of $\theta^\mu(\cdot)$ to $\theta(\cdot)$, the Skorohod representation, and using $\mu \ell n_\mu \rightarrow u$ and (C1)–(C2), it can then be shown

$$\begin{aligned} & \sum_{\tilde{\theta}=1}^{\Theta} \frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} b_i^\gamma(\tilde{\mathbf{f}}^\mu(\mu \ell n_\mu)) \mathbb{E}_{\ell n_\mu} f_k(i) I_{\{\theta^\mu(\mu \ell n_\mu)=\tilde{\theta}\}} \\ & \rightarrow b_i^\gamma(\hat{\mathbf{f}}(u) + \mathbf{F}(\theta(u))) f(i, \theta(u)) \quad \text{in probability as } \mu \rightarrow 0. \end{aligned} \quad (56)$$

Using a similar argument for the second term in (55), we conclude that, as $\mu \rightarrow 0$,

$$\begin{aligned} & \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \quad \times \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla'_r y \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} [f_k(s(k)) - F_{\min}(\theta(\ell n_\mu))] \right] \right] \\ & \rightarrow \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \quad \times \left[\int_t^{t+u} \nabla'_r y(\mathbf{X}(v), \theta(v)) \right. \\ & \quad \times \left. \left[\mathbf{b}^\gamma(\hat{\mathbf{f}}(v) + \mathbf{F}(\theta(v))) \cdot \mathbf{F}(\theta(v)) - F_{\min}(\theta(v)) \right] dv \right]. \end{aligned} \quad (57)$$

Finally, similar to (54),

$$\begin{aligned} & \mathbb{E}h(\mathbf{X}^\mu(t_\iota), \theta^\mu(t_\iota) : \iota \leq \kappa_0) \\ & \quad \times \left[\sum_{\ell: \ell\delta_\mu=t}^{t+u} \delta_\mu \nabla'_r y \left[\frac{1}{n_\mu} \sum_{k=\ell n_\mu}^{\ell n_\mu + n_\mu - 1} r(k) \right] \right] \\ & \rightarrow \mathbb{E}h(\mathbf{X}(t_\iota), \theta(t_\iota) : \iota \leq \kappa_0) \\ & \quad \times \left[\int_t^{t+u} \nabla'_r y(\mathbf{X}(v), \theta(v)) r(v) dv \right] \quad \text{as } \mu \rightarrow 0. \end{aligned} \quad (58)$$

Combining the above two steps concludes the proof.

APPENDIX B PROOF OF THEOREM 4.2

We first prove that each subsystem (the ODE (26) associated with each $\bar{\theta} \in \mathcal{Q}$ when $\theta(t) = \bar{\theta}$ is held fixed) is globally asymptotically stable $\mathbb{R}_{[0, \eta)}$ is its global attracting set. Define the Lyapunov function:

$$V_{\bar{\theta}}(r) = r^2.$$

Taking the time derivative, and applying (26), we obtain

$$\frac{d}{dt} V_{\bar{\theta}}(r) = 2r \cdot [\mathbf{b}^\gamma(\mathbf{F}(\bar{\theta})) \cdot \mathbf{F}(\bar{\theta}) - F_{\min}(\bar{\theta}) - r]$$

Since the objective function value at various states is bounded for each $\bar{\theta} \in \mathcal{Q}$,

$$\frac{d}{dt} V_{\bar{\theta}}(r) \leq 2r \cdot [C(\gamma, \bar{\theta}) - r]$$

for some constant $C(\gamma, \bar{\theta})$. Recall the smooth best-response sampling strategy $\mathbf{b}^\gamma(\cdot)$ in Definition 3.1. The parameter γ simply determines the magnitude of perturbations applied to the objective function. It is then clear that $C(\gamma, \bar{\theta})$ is monotonically increasing in γ .

In view of (59), for each $\eta > 0$, $\hat{\gamma}$ can be chosen small enough such that, if $\gamma \leq \hat{\gamma}$ and $r \geq \eta$,

$$\frac{d}{dt} V_{\bar{\theta}}(r) \leq -V_{\bar{\theta}}(r)$$

Therefore, each subsystem is globally asymptotically stable and, for $\gamma \leq \hat{\gamma}$,

$$\lim_{t \rightarrow \infty} d(r(t), \mathbb{R}_{[0, \eta)}) = 0.$$

Finally, stability of the regime-switching ODE (25) is examined. We can use the above Lyapunov function to extend [14, Corollary 12] to prove global asymptotic stability w.p.1.

Theorem B.1 ([14, Corollary 12]): Consider the switching system (27) in Definition 4.1, where $\theta(t)$ is the state of a continuous time Markov chain with generator Q . Define $\bar{q} := \max_{\theta \in \mathcal{Q}} |q_{\theta\theta}|$ and $\tilde{q} := \max_{\theta, \theta' \in \mathcal{Q}} q_{\theta\theta'}$. Suppose there exist continuously differentiable functions $V_\theta : \mathbb{R}^n \rightarrow \mathbb{R}^+$, $\theta \in \mathcal{Q}$, strictly increasing functions $a_1, a_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ with $a_1(0) = a_2(0) = 0$ and $a_1(t), a_2(t) \rightarrow \infty$ as $t \rightarrow \infty$, a real number $v > 1$ such that the following hold:

$$1) \quad a_1(d(Y, \mathcal{H})) \leq V_\theta(Y) \leq a_2(d(Y, \mathcal{H})), \quad \forall Y \in \mathbb{R}^r, \theta \in \mathcal{Q},$$

- 2) $\frac{\partial V_\theta}{\partial X} f_\theta(Y) \leq -\lambda V_\theta(Y)$, $\forall Y \in \mathbb{R}^r$, $\forall \theta \in \mathcal{Q}$,
- 3) $V_\theta(Y) \leq v V_{\theta'}(Y)$, $\forall Y \in \mathbb{R}^r$, $\theta, \theta' \in \mathcal{Q}$,
- 4) $(\lambda + \tilde{q})/\bar{q} > v$.

Then, the regime-switching system (27) is globally asymptotically stable almost surely.

The quadratic Lyapunov functions (59) satisfies Hypothesis 2) in Theorem B.1; see (59). Further, since the Lyapunov functions are the same for all subsystems $\theta \in \mathcal{Q}$, existence of $v > 1$ in Hypothesis 3) is automatically guaranteed. Hypothesis 4) simply ensures that the switching signal $\theta(t)$ is slow enough. Given that $\lambda = 1$ in hypothesis 2), it remains to ensure that the generator Q of Markov chain $\theta(t)$ satisfies $1 + \tilde{q} > \bar{q}$. This is satisfied since $|q_{\theta\theta'}| \leq 1$ for all $\theta, \theta' \in \mathcal{Q}$; see (4).

REFERENCES

- [1] V. I. Norkin, Y. M. Ermoliev, and A. Ruszczyński, "On optimal allocation of indivisibles under uncertainty," *Oper. Res.*, vol. 46, no. 3, pp. 381–395, 1998.
- [2] J. R. Swisher, P. D. Hyden, S. H. Jacobson, and L. W. Schruben, "A survey of simulation optimization techniques and procedures," in *Proc. 2000 Winter Simulation Conf.*, vol. 1, 2000, pp. 119–128.
- [3] K. Park and Y. Lee, "An on-line simulation approach to search efficient values of decision variables in stochastic systems," *Int. J. Adv. Manuf. Technol.*, vol. 25, no. 11–12, pp. 1232–1240, 2005.
- [4] V. Krishnamurthy, X. Wang, and G. Yin, "Spreading code optimization and adaptation in CDMA via discrete stochastic approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1927–1949, Sep. 2004.
- [5] I. Berenguer, X. Wang, and V. Krishnamurthy, "Adaptive MIMO antenna selection via discrete stochastic optimization," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4315–4329, Nov. 2005.
- [6] G. C. Pflug, *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*. Norwell, MA: Kluwer Academic Publishers, 1996.
- [7] A. Benveniste, M. Metivier, and P. Prioret, *Adaptive Algorithms and Stochastic Approximations*. New York: NY: Springer-Verlag, 1990.
- [8] D. Yan and H. Mukai, "Stochastic discrete optimization," *SIAM J. Control Optim.*, vol. 30, no. 3, pp. 594–612, May 1992.
- [9] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. New York, NY: Springer-Verlag, 2003.
- [10] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. MIT Press, 1998, vol. 2.
- [11] G. Yin and Q. Zhang, *Continuous-time Markov Chains and Applications: A Singular Perturbation Approach*. New York: Springer Verlag, 1998.
- [12] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*. Springer Verlag, 2009, vol. 63.
- [13] D. Chatterjee and D. Liberzon, "Stability analysis of deterministic and stochastic switched systems via a comparison principle and multiple lyapunov functions," *SIAM J. Control Optim.*, vol. 45, no. 1, pp. 174–206, 2007.
- [14] —, "On stability of randomly switched nonlinear systems," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2390–2394, Dec. 2007.
- [15] S. Andradóttir, "An overview of simulation optimization via random search," *Handbooks in Operations Research and Management Science*, vol. 13, pp. 617–631, 2006.
- [16] R. Y. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. Chichester, England: Wiley, 1993.
- [17] H. Chen and B. W. Schmeiser, "Stochastic root finding via retrospective approximation," *IIE Transactions*, vol. 33, no. 3, pp. 259–275, Mar. 2001.
- [18] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello, "The sample average approximation method for stochastic discrete optimization," *SIAM J. Optim.*, vol. 12, no. 2, pp. 479–502, 2002.
- [19] T. Homem-De-Mello, "Variable-sample methods for stochastic optimization," *ACM Trans. Model. Comput. Sim.*, vol. 13, no. 2, pp. 108–133, Apr. 2003.
- [20] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ: Wiley, 2003.
- [21] W. J. Gutjahr and G. C. Pflug, "Simulated annealing for noisy cost functions," *Journal of Global Optimization*, vol. 8, no. 1, pp. 1–13, Jan. 1996.
- [22] A. A. Prudius and S. Andradóttir, "Averaging frameworks for simulation optimization with applications to simulated annealing," *Naval Research Logistics*, vol. 59, no. 6, pp. 411–429, Sep. 2012.
- [23] F. Glover and M. Laguna, "Tabu search," in *Encyclopedia of Operations Research and Management Science*, S. I. Gass and C. M. Harris, Eds. Springer, 1996, pp. 671–701.
- [24] M. H. Alrefaei and S. Andradóttir, "Discrete stochastic optimization using variants of the stochastic ruler method," *Naval Research Logistics*, vol. 52, no. 4, pp. 344–360, Jun. 2005.
- [25] S. Andradóttir, "A global search method for discrete stochastic optimization," *SIAM J. Optim.*, vol. 6, no. 2, pp. 513–530, May 1996.
- [26] —, "Accelerating the convergence of random search methods for discrete stochastic optimization," *ACM Trans. Model. Comput. Sim.*, vol. 9, no. 4, pp. 349–380, Oct. 1999.
- [27] G. Yin, V. Krishnamurthy, and C. Ion, "Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization," *SIAM J. Optim.*, vol. 14, no. 4, pp. 1187–1215, 2004.
- [28] S. Andradóttir and A. A. Prudius, "Balanced explorative and exploitative search with estimation for simulation optimization," *INFORMS J. Comput.*, vol. 21, no. 2, pp. 193–208, Spring 2009.
- [29] V. I. Norkin, G. C. Pflug, and A. Ruszczyński, "A branch and bound method for stochastic global optimization," *Mathematical programming*, vol. 83, no. 1–3, pp. 425–450, 1998.
- [30] L. Shi and S. Ólafsson, "Nested partitions method for global optimization," *Oper. Res.*, vol. 48, no. 3, pp. 390–407, 2000.
- [31] L. J. Hong and B. L. Nelson, "Discrete optimization via simulation using compass," *Oper. Res.*, vol. 54, no. 1, pp. 115–129, 2006.
- [32] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2–3, pp. 235–256, 2002.
- [33] J.-Y. Audibert, S. Bubeck, and R. Munos, "Best arm identification in multi-armed bandits," in *Proc. 23th Conf. Learning Theory*, Haifa, Israel, Jun. 2010, pp. 41–53.
- [34] A. Garivier and E. Moulines, "On upper-confidence bound policies for non-stationary bandit problems," *arXiv:0805.3415 [math.ST]*.
- [35] S. Yakowitz, P. L'ecuyer, and F. Vázquez-Abad, "Global stochastic optimization with low-dispersion point sets," *Oper. Res.*, vol. 48, no. 6, pp. 939–950, 2000.
- [36] J. Hofbauer and W. H. Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, vol. 70, no. 6, pp. 2265–2294, Nov. 2002.
- [37] M. Benaïm, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions, Part II: Applications," *Math. Oper. Res.*, vol. 31, no. 4, pp. 673–695, Nov. 2006.
- [38] D. Fudenberg and D. K. Levine, "Conditional universal consistency," *Games Econ. Behav.*, vol. 29, no. 1–2, pp. 104–130, Oct. 1999.
- [39] D. Fudenberg and D. Levine, "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control*, vol. 19, no. 5–7, pp. 1065–1089, 1995.
- [40] J. Hannan, "Approximation to bayes risk in repeated play," *Contributions to the Theory of Games*, vol. 3, pp. 97–139, 1957.
- [41] M. Benaïm and M. Faure, "Consistency of vanishingly smooth fictitious play," *Math. Oper. Res.*, 2012.
- [42] P. Billingsley, *Convergence of probability measures*. New York: Wiley, 1968.
- [43] D. Liberzon, *Switching in Systems and Control*. Springer, 2003.
- [44] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes With Application to Stochastic Systems Theory*. Cambridge, MA: MIT Press, 1984.